

University of Groningen

## Extensions of graphical models with applications in genetics and genomics

Behrouzi, Pariya

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Behrouzi, P. (2018). *Extensions of graphical models with applications in genetics and genomics*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# **Extensions of Graphical Models with Applications in Genetics and Genomics**

Pariya Behrouzi





university of  
 groningen

# **Extensions of Graphical Models with Applications in Genetics and Genomics**

**PhD thesis**

to obtain the degree of PhD at the  
University of Groningen  
on the authority of the  
Rector Magnificus Prof. E. Sterken  
and in accordance with  
the decision by the College of Deans.

This thesis will be defended in public on

Friday 19 January 2018 at 12.45 hours

by

**Pariya Behrouzi**

born on 11 June 1986  
in Zanjan, Iran

**Supervisor**

Prof. E. C. Wit

**Assessment Committee**

Prof. K. Strimmer

Prof. C. H. Gräfin zu Eulenburg

Prof. E. R. van den Heuvel





*To my parents*

*&*

*To Reza*





# Contents

<b>Contents</b>	<b>ix</b>
<b>Chapter 1: General Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Some basic genetics . . . . .	2
1.2.1 Probabilistic model of meiosis . . . . .	4
1.2.2 Genetic map . . . . .	4
1.2.3 Genetic linkage study . . . . .	5
1.3 Graphical models . . . . .	6
1.3.1 Directed acyclic graphical models . . . . .	7
1.3.2 Undirected graphical models . . . . .	8
1.3.3 Chain graph models . . . . .	10
1.4 Gaussian Copula . . . . .	11
1.4.1 Dependence in Gaussian copula . . . . .	12
1.5 Outline of thesis contribution . . . . .	14
<b>References</b>	<b>17</b>
<b>Chapter 2: Detecting Epistatic Selection with Partially Observed Genotype Data Using Copula Graphical Models</b>	<b>19</b>
2.1 Introduction . . . . .	20
2.2 Genetic background of epistatic selection . . . . .	22
2.2.1 Meiosis . . . . .	22
2.2.2 Recombinant Inbred Lines . . . . .	23
2.2.3 Genome-wide association study . . . . .	23
2.2.4 Epistatic phenotype . . . . .	24
2.3 Graphical model for epistatic selection . . . . .	25

2.3.1	Gaussian copula graphical model . . . . .	26
2.3.2	$\ell_1$ penalized inference of Gaussian copula graphical model . . . . .	27
2.3.3	Selection of the tuning parameter . . . . .	30
2.3.4	Inference uncertainty . . . . .	32
2.4	Simulation study . . . . .	33
2.5	Detecting genomic signatures of epistatic selection . . . . .	35
2.5.1	Epistatic selection in <i>Arabidopsis thaliana</i> . . . . .	35
2.5.2	Genetic inbreeding experiment in maize . . . . .	39
2.6	Discussion . . . . .	40
2.7	Supplementary Materials . . . . .	42
<b>References</b>		<b>45</b>
 <b>Chapter 3: De novo construction of q-ploid linkage maps using discrete graphical models</b>		
		<b>49</b>
3.1	Introduction . . . . .	50
3.2	Genetic background on linkage map . . . . .	52
3.2.1	Linkage map for diploids and polyploids . . . . .	52
3.2.2	Mapping population . . . . .	53
3.2.3	Meiosis and Markov dependence . . . . .	54
3.3	Algorithm to detect linkage map . . . . .	58
3.3.1	Estimating marker-marker network . . . . .	58
3.3.2	Determining linkage groups . . . . .	60
3.3.3	Ordering markers . . . . .	61
3.4	Simulation study . . . . .	62
3.4.1	Diploid species . . . . .	63
3.4.2	Polyploid species . . . . .	66
3.5	Construction of linkage map for diploid barley . . . . .	66
3.6	Construction of linkage map for tetraploid potato . . . . .	68
3.7	Conclusion . . . . .	69
3.8	Supporting information . . . . .	71
<b>References</b>		<b>79</b>

<b>Chapter 4: netgwas: An R Package for Network-Based Genome Wide Association Studies</b>	<b>81</b>
4.1 Introduction . . . . .	82
4.2 Methodological background . . . . .	83
4.3 Package netgwas . . . . .	87
4.3.1 User interface . . . . .	87
4.3.2 netmap . . . . .	91
4.3.3 netsnp . . . . .	93
4.3.4 netphenogeno . . . . .	94
4.4 Discussion . . . . .	100
<b>References</b>	<b>103</b>
<b>Chapter 5: Dynamic Chain Graph Models for Ordinal Time Series Data</b>	<b>107</b>
5.1 Introduction . . . . .	107
5.2 Methods . . . . .	109
5.2.1 Dynamic chain graph models . . . . .	109
5.2.2 Gaussian Copula . . . . .	110
5.2.3 Model definition . . . . .	110
5.2.4 Penalized EM inference . . . . .	112
5.2.5 Selection of tuning parameters . . . . .	118
5.3 Simulation study . . . . .	120
5.4 Netherlands Study of Depression and Anxiety . . . . .	121
5.5 Discussion . . . . .	123
5.6 Appendix . . . . .	123
<b>References</b>	<b>127</b>
<b>Chapter 6: Conclusions</b>	<b>131</b>
6.1 General overview of thesis . . . . .	131
6.2 Highlight of the results . . . . .	132
6.3 Discussion . . . . .	133
6.3.1 Gaussian copula . . . . .	134
6.3.2 Ordering markers . . . . .	134
6.3.3 Interpretation of multi-trait networks . . . . .	134
6.3.4 Granger causality . . . . .	136

---

6.4	Future Work . . . . .	136
6.4.1	Epistatic interactions network . . . . .	137
6.4.2	Linkage map . . . . .	137
6.4.3	Directed graphs for mixed discrete-continuous data . . . . .	137
6.4.4	Nonlinear dynamic time-series network . . . . .	138
6.4.5	Network inference and modeling networks data . . . . .	138
	<b>Summary</b>	<b>139</b>
	<b>Samenvatting</b>	<b>141</b>